Geophysical Prospecting

EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Geophysical Prospecting, 2015, **63**, 35–54 doi: 10.1111/1365-2478.12163

Wavefield tomography based on local image correlations

Francesco Perrone[1]*, Paul Sava[2] and Jacopo Panizzardi[3]

[1]*Now at CGG, Crompton Way, Crawley, West Sussex, RH10 9QN, UK,* [2]*Center for Wave Phenomena, Department of Geophysics, Colorado School of Mines, 1500 Illinois Street, Golden Colorado 80401, USA, and* [3]*ENI E&P, Via Emilia, 1 20097 San Donato Milanese (MI), Italy*

Received March 2013, revision accepted February 2014

The estimation of a velocity model from seismic data is a crucial step for obtaining a high-quality image of the subsurface. Velocity estimation is usually formulated as an optimization problem where an objective function measures the mismatch between synthetic and recorded wavefields and its gradient is used to update the model. The objective function can be defined in the data-space (as in full-waveform inversion) or in the image space (as in migration velocity analysis). In general, the latter leads to smooth objective functions, which are monomodal in a wider basin about the global minimum compared to the objective functions defined in the data-space. Nonetheless, migration velocity analysis requires construction of common-image gathers at fixed spatial locations and subsampling of the image in order to assess the consistency between the trial velocity model and the observed data. We present an objective function that extracts the velocity error information directly in the image domain without analysing the information in common-image gathers. In order to include the full complexity of the wavefield in the velocity estimation algorithm, we consider a two-way (as opposed to one-way) wave operator, we do not linearize the imaging operator with respect to the model parameters (as in linearized wave-equation migration velocity analysis) and compute the gradient of the objective function using the adjoint-state method. We illustrate our methodology with a few synthetic examples and test it on a real 2D marine streamer data set.

Key words: Tomography, Wavefield, Adjoint-state method.

INTRODUCTION

Seismic imaging involves the estimation of wave propagation velocities in the subsurface from seismic data recorded at the surface. Seismic velocities are related to other physical parameters (for example, density and compressibility, which characterize the lithology of the Earth) and rock-mechanics parameters (for example, porosity and fluid overpressure, which are crucial in reservoir engineering) (Carcione 2007).

Wave-equation tomography (Tarantola 1984; Woodward 1992; Biondi and Sava 1999) is a family of techniques that estimate the velocity model parameters from finite bandwidth signals recorded at the surface. The inversion is usually formulated as an optimization problem where the correct velocity model minimizes an objective function that measures the inconsistency between a trial model and the observations. The objective function can be defined either in the data space (full-waveform inversion) or in the image-space (migration velocity analysis).

Full-waveform inversion (FWI) (Tarantola 1984; Pratt 1999; Sirgue and Pratt 2004) addresses the estimation problem in the data space and measures the mismatch between the observations and simulated data. Full-waveform inversion reconstructs a model that explains the data in the sense that the mismatch between acquired and simulated waveforms is minimized. By matching both traveltimes and amplitudes, full-waveform inversion allows one to achieve high resolution (Sirgue *et al.* 2010). Nonetheless, a source estimate is needed,

*E-mail: fperrone@mines.edu

© 2014 European Association of Geoscientists & Engineers

the physics of wave propagation (for example, isotropic versus anisotropic, acoustic versus elastic, etc.) must be correctly modelled and a good parametrization (for example, impedance versus velocity contrasts) is crucial (Kelly *et al.* 2010). Moreover, an accurate initial model is key to avoid cycle skipping and converge to the global minimum (instead of a local minimum) of the objective function.

Because of the non-linearity of the wavefields with respect to the velocity model, the objective function in the data domain is highly multimodal (Santosa and Symes 1989) and local optimization methods can easily get trapped into local minima and fail to converge to the correct model. This is particularly true for reflection full-waveform inversion. Refraction full-waveform inversion focuses on diving waves and mutes the data, retaining only the diving energy (Pratt 1999). This leads to a better-behaved objective function but requires very long offsets in order to record the refracted energy. Moreover, this approach limits the depth at which a robust inversion result can be expected.

Migration velocity analysis (MVA) (Fowler 1985; Faye and Jeannot 1986; Al-Yahya 1989; Chavent and Jacewitz 1995; Biondi and Sava 1999; Sava, Biondi and Etgen 2005; Albertin *et al.* 2006) defines the objective function in the image space and is based on the semblance principle (Sattlegger 1975; Al-Yahya 1989). If the velocity model is correct, images from different experiments must be consistent with each other because a single Earth model generates the recorded data. A measure of consistency is usually computed through conventional semblance (Taner and Koehler 1969) or differential semblance (Symes and Carazzone 1991; Symes 1993). These two functionals analyse a set of migrated images at fixed locations in space; they consider all the shots that illuminate the points under investigation. Migration velocity analysis leads to smooth objective functions and well-behaved optimization problems (Symes 1991; Symes and Carazzone 1991) and is less sensitive than full-waveform inversion to the initial model. On the other hand, because we do not use amplitudes in the imaging step, the estimated model has lower resolution than the ideal full-waveform inversion result (Uwe Albertin, pers. comm.).

Migration velocity analysis measures either the invariance of the migrated images in an auxiliary dimension (reflection angle, shot, etc.) (Al-Yahya 1989; Rickett and Sava 2002; Sava and Fomel 2003; Xie and Yang 2008) or focusing in an extended space (Rickett and Sava 2002; Symes 2008; Sava and Vasconcelos 2009; Yang and Sava 2011). Migration can be implemented with one-way or two-way extrapolation engines, thus we can use either downward continuation (Stolt 1978; Gazdag and Sguazzero 1984; Stoffa *et al.* 1990) or reverse-time migration (Baysal, Kosloff, and Sherwood 1983; McMechan 1983) in the wavefield tomography procedure. All these approaches require the migration of the entire survey in order to analyse the moveout curve in common-image gathers or measure focusing at a specific spatial location. The dimensionality of the (extended) image space and computational complexity of the velocity analysis step rapidly explodes for realistic case scenarios. Moreover, because of the high memory requirement for storing the partial information from each experiment, only a subset of the image points can be considered in the evaluation of the objective function. Illumination holes and/or irregular acquisition geometries can also impact the quality of the common-image gathers but no systematic study of this problem is reported in the literature to our knowledge.

Reverse-time migration (Baysal *et al.* 1983; McMechan 1983) is routinely used in exploration geophysics because of its ability to correctly handle the full complexity of the wave propagation phenomena (under the assumption that the physics of wave propagation in the subsurface is correctly modelled). Its computational cost is nevertheless still prohibitive and limits its integration into a migration velocity analysis loop that requires the extraction of common-image gathers for moveout analysis. A migration velocity analysis procedure based on a full-wave propagation engine allows exploitation of more complex wave phenomena (e.g., overturning reflection, prismatic waves and multiples) and increases the amount of information in the data that can be used (Farmer *et al.* 2006).

Reverse-time migration works only in common-shot and common-receiver configurations, so the choice of the shot domain for performing migration velocity analysis with a two-way engine may be preferable. Here, we use the word 'shot' in its broader sense to include synthetic shot-gathers like plane-wave sources (Whitmore, 1995; Zhang *et al.* 2005; Liu *et al.* 2006; Stoffa *et al.* 2006), random shot-encoded sources (Morton and Ober 1998; Romero *et al.* 2000) or any other phase-/amplitude-encoded source (Soubaras 2006; Perrone and Sava 2012a) in the range of shot-profile migration. Reverse-time migration automatically and without additional costs outputs shot-domain common-image gathers; unfortunately, the migrated-shot domain is notoriously not optimal for velocity estimation because of migration artefacts (Stolk and Symes 2004) and noisy common-image gathers (Zhang *et al.* 2010). Even when the velocity model is correct, the reflection events in shot domain image gathers are very noisy and semblance is challenging to evaluate. Previously

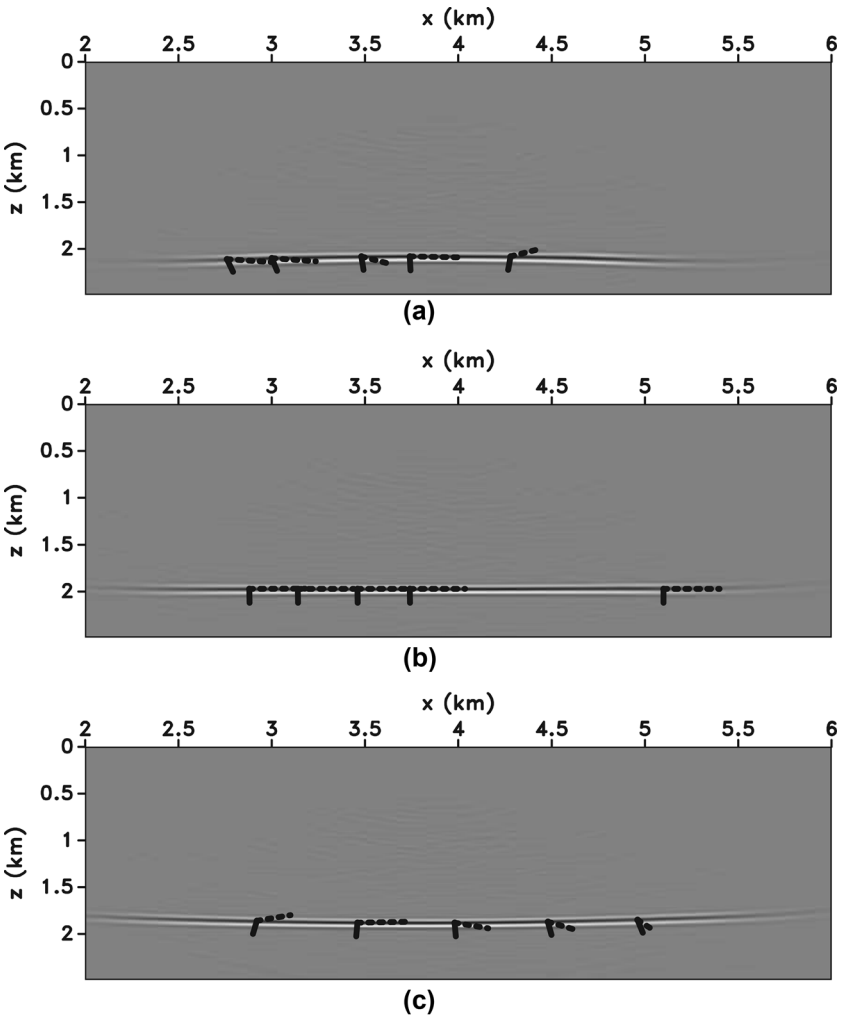


Figure 1 Relationship between the dip field and the apparent displacement: dip (solid) and displacement (dashed) vector fields for different values of slowness error for single horizontal reflectors: (a) $\frac{\Delta m}{m} = 10\%$, $\frac{\Delta m}{m} = 0$ and (c) $\frac{\Delta m}{m} = 10\%$. The correct slowness value is 0.5 s/km.

proposed techniques in the shotdomain consider all shots in the survey, thus exploiting the power of stacking to attenuate the effects of noise and artefacts in the common-image gathers (de Vries and Berkhout 1984; Yilmaz and Chambers 1984; Al-Yahya 1989; Chavent and Jacewitz 1995; Xie and Yang 2008).

We propose an objective function that evaluates the degree of semblance between images through local correlations in the image space and does not require to pick moveout or evaluate focusing in common-image gathers (CIGs). We use the morphologic relationship between images from nearby experiments to define an objective function that measures shifts in the image space (Perrone and Sava 2012b). The methodology described in this work takes advantage of the local coherency of the migrated reflectors in the image domain to overcome the limitations of shot-domain CIGs. We compute the gradient of the objective function with the adjoint-state method (Plessix 2006). Our approach does not require to store partial information about common-image gathers and avoids the need to pick moveout on gathers and allows us to include all points illuminated by the seismic experiments in the velocity model building.

THEORY

We define a measure of velocity error, $r(\mathbf{x})$, where $\mathbf{x} = (z, x)$, in the image domain using a penalized local correlation that measures the apparent shifts between single shots migrated images. We define the objective function of the optimization problem as the energy (l_2 norm) of $r(\mathbf{x})$. Other norms (l_1, Huber norm, hybrid, etc.) can be used to promote sparsity and gain robustness against non-Gaussian noise in the data (Brossier, Operto and Virieux 2010). We use adjoint-state calculations to compute the gradient of the objective function.

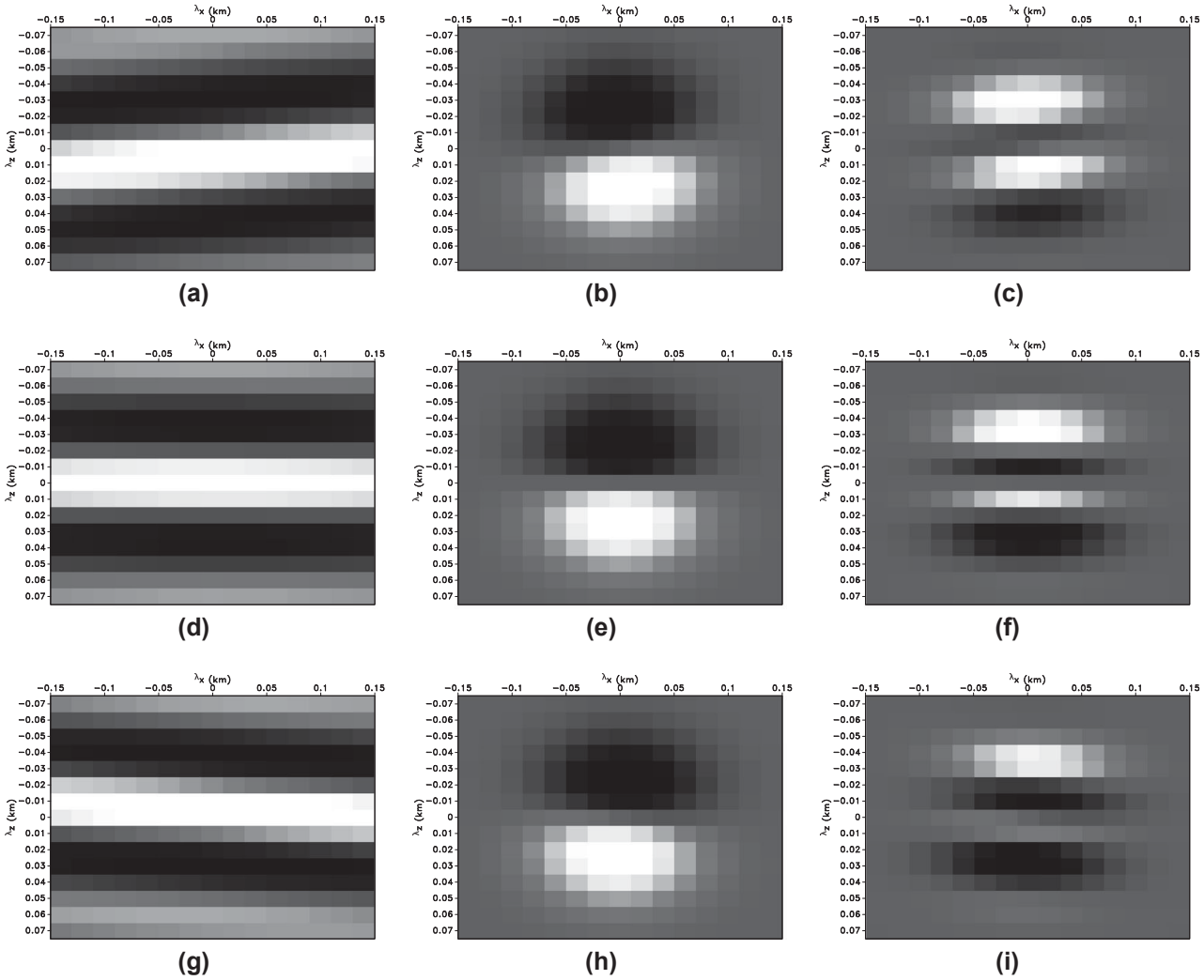


Figure 2 Measure of the relative displacement between two images using penalized local correlations. We consider 2 images from nearby experiments of the reflector in Fig. 1; each row (from top to bottom) represents a different model: too high, correct and too low; the columns show (from left to right) the local correlation panel, the penalty operator used and the penalized local correlation. For each model, we pick a point on the reflector with lateral coordinate $x = 3.5$ km. The vertical coordinate changes as a function of model parameters because the depth of the reflector changes. Observe the asymmetry of the penalized local correlations for the wrong models: the mean value (i.e., the stack over the correlation lags) of the penalized correlations gives us a measure of the relative shift between the two images.

Image correlation objective function

There are at least two options for measuring similarity between two images. We can compute the point-wise difference function as:

$$r(\mathbf{x}) = R_{i+1}(\mathbf{x}) - R_i(\mathbf{x}), \tag{1}$$

where $R_i(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$ are the migrated images for ith and $(i+1)$th shot, respectively. The two shots are assumed to be close and to image the same area in the subsurface. Plessix (2006) and Symes (1993) used the energy of the image difference as a regularization term for full-waveform inversion and not as a stand-alone objective function. Notice that if the velocity model used in the migration is severely inaccurate, we can have cycle skipping in the image space or, in other words, the difference between the two images can produce two events if the position of the same reflector in the two images changes more than a quarter of a wavelength. This problem is analogous to what happens in full-waveform inversion in the data space.

An alternative solution is to restate the semblance principle as follows: if the velocity model is correct, two images from

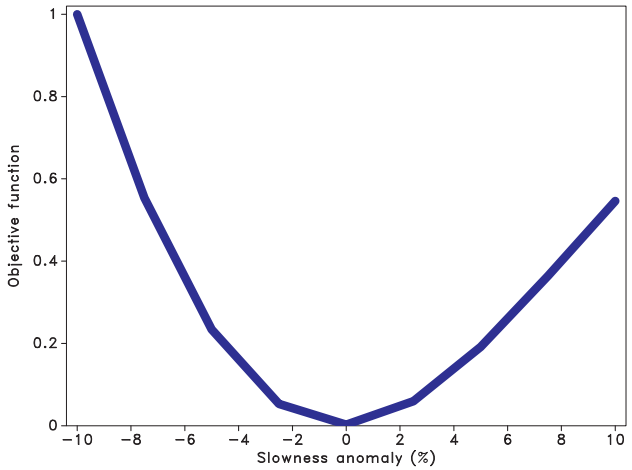


Figure 3 Values of the objective function for different errors in the slowness model. We consider constant perturbations ranging from −10% to 10% of the exact value (0.5 s/km).

nearby experiments construct the final image along the direction of the reflectors (Perrone and Sava 2012). This is equivalent to the standard assumption that, if the model is correct, the prestack image cube is invariant with respect to the shot position, i.e., no moveout in the shot-domain common-image gathers (Xie and Yang 2008).

The structural dip is an attribute that is commonly extracted in seismic processing (van Vliet and Verbeek 1995; Fomel 2002). The similarity of two images perpendicular to the structural dip can be measured by means of penalized local correlations (Perrone and Sava 2012). We illustrate this idea on a simple model with a single horizontal density interface; the correct velocity model is constant and equal to 2.0 km/s. Figure 1 shows the dip and apparent displacement vector field (solid and dashed arrows, respectively) at particular spatial locations. Observe that the two vectors are orthogonal when the velocity model is correct.

The computation of the displacement field requires a highly non-linear procedure, which extracts the maximum of the local correlations at every image point (Hale, Cox and Hatchell 2008). We reformulate the problem of evaluating the orthogonality between dip and displacement using local correlations between images $R_i(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$ (Hale 2006):

$$c_i(\mathbf{x}, \boldsymbol{\lambda}) = \int w(\mathbf{x} - \boldsymbol{\xi})\, R_i\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) d\boldsymbol{\xi}, \qquad (2)$$

penalized along the dip direction. The 2D vector $\boldsymbol{\lambda} = (\lambda_z, \lambda_x)$ denotes the correlation lag in the image space and the index i scans the shot position. The range of λ is chosen according to the wavelength of the signal. As a rule of thumb the size of the local window is approximately equal to the maximum wavelength of the signal. Local Gaussian windows $w(\mathbf{x} - \boldsymbol{\xi}) = \exp(-1/2((z - \xi_z)^2/\sigma_z^2 + (x - \xi_x)^2/\sigma_x^2))$ centred about the image points $\boldsymbol{x}$ weight the input images and allow computation of local cross-correlations. $\boldsymbol{\xi} = (\xi_z, \xi_x)$ represents the dummy variable of integration in the correlation and it is defined in the same space as $\boldsymbol{x}$. If the velocity model is correct, the maximum of the correlation lies along the reflector slope (Fig. 1b); otherwise, we observe a deviation that represents the relative shift of one image with respect to the other (Fig. 1a,c). This criterion measures the similarity of the structures in the image domain and indirectly measures the flatness of shot-domain common-image gathers. Since we do not constrain the vertical scale and P-wave propagation velocities depend on three Thomsen's parameters V_{p0}, δ and ϵ (Thomsen 1986), this approach cannot resolve anisotropy. In presence of anisotropy, it is indeed possible to obtain perfectly focused images (i.e., flat common-image gathers) with incorrect models because of the trade-off between the different parameters. To measure the shift between the two images, we consider the penalty operator:

$$P(\mathbf{x}, \boldsymbol{\lambda}) = \boldsymbol{\lambda} \cdot \nu(\mathbf{x}), \qquad (3)$$

where λ is the correlation lag vector and the dip vector $\nu(\mathbf{x})$ is the normal to the reflector at $\boldsymbol{x}$. The dip is computed from either one of the images $R_i(\mathbf{x})$ or $R_{i+1}(\mathbf{x})$; we used always the first image of the pair of shots considered in the correlation. The penalty operator $P(\mathbf{x}, \boldsymbol{\lambda})$ is a linear function in the dip direction (normal to the reflector) and the isopenalty lines are parallel to the reflector. The penalty operator is identically zero along the reflector.

Figure 2 shows the local correlations, penalty operators and penalized local correlations for a particular location on the reflector in the three cases in Fig. 1. The rows of Fig. 2. refer to different models (from top to bottom: a too high, the correct, and a too low velocity model). The point on the image has lateral position $x = 3.5$ km and vertical that depends on the imaged reflector. Figure 2(a,d,g) shows the local correlations computed from two nearby shots; observe the shift in the peak of the local correlation and the orientation of the correlation panels. Figure 2(b,e,h) shows the penalty operators that we use to highlight the inconsistency between the dip measured on the image and the local correlations. The orientation of the penalty operators does indeed depend on the measured dip. We applied a Gaussian taper to the penalty operator in order to have a smooth transition to zero along

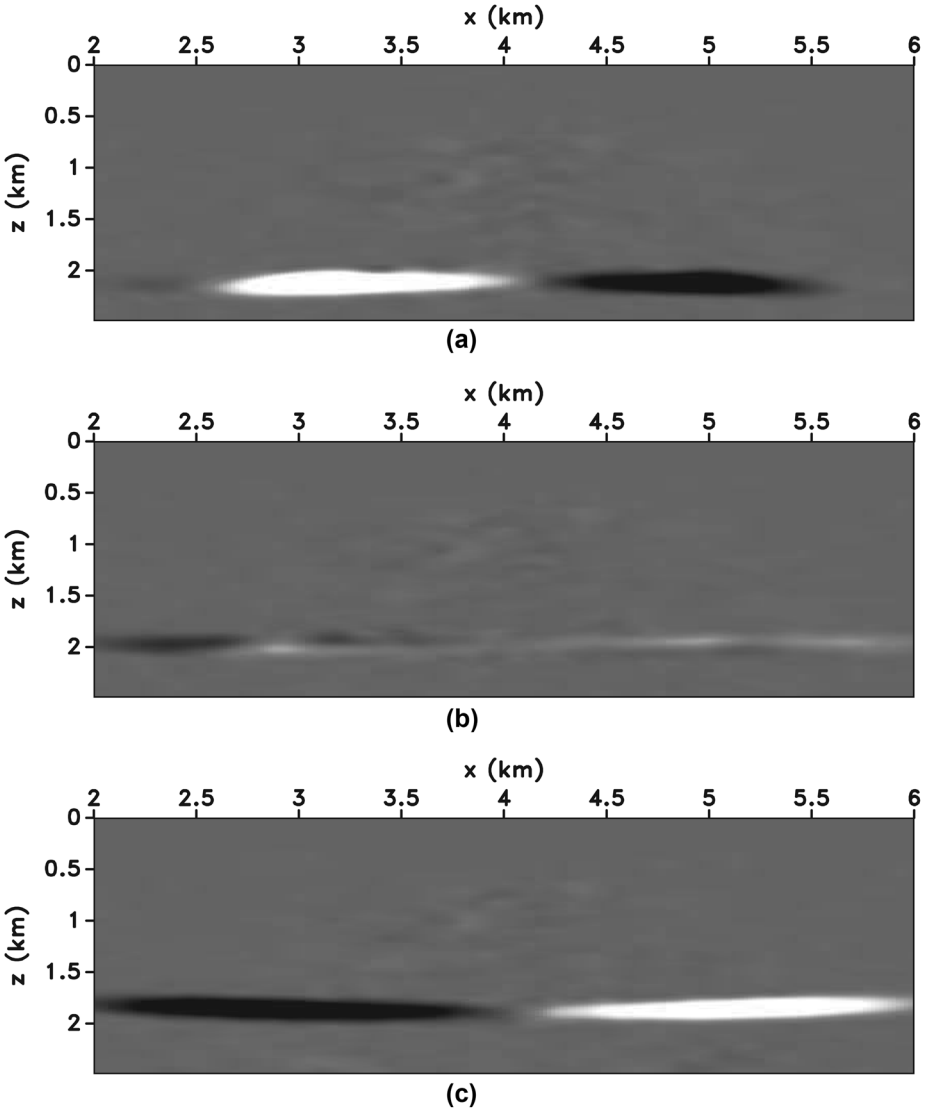


Figure 4 For the three cases in Fig. 1, we compute the shifts $r(\mathbf{x})$ according to equation (4). The sign of the shifts indicates the relative displacement of one image with respect to the other.

the sides of the penalty operator panel and thus remove possible artefacts due to abrupt truncations. Figure 2(c,f,i) shows the penalized correlations. The apparent shift is now clear as a sign unbalance in the panels. When the velocity model is correct, the penalized correlation is an odd function (i.e., zero mean value) (Fig. 2f); on the other hand, the panels lose their symmetry when the model is incorrect (Fig. 2c,i). The sign of the mean value of the penalized panels measures the apparent shift between the two images.

We define the velocity error measure as the average value of the local correlations $c_i(\mathbf{x}, \lambda)$ penalized by the penalty operator $P(\mathbf{x}, \lambda)$:

$$r_i(\mathbf{x}) = \int P(\mathbf{x}, \boldsymbol{\lambda})\, c_i(\mathbf{x}, \boldsymbol{\lambda})\, d\boldsymbol{\lambda}. \tag{4}$$

This quantity measures the relative shift of one image with respect to the other directly in the image space and can be used for velocity analysis.

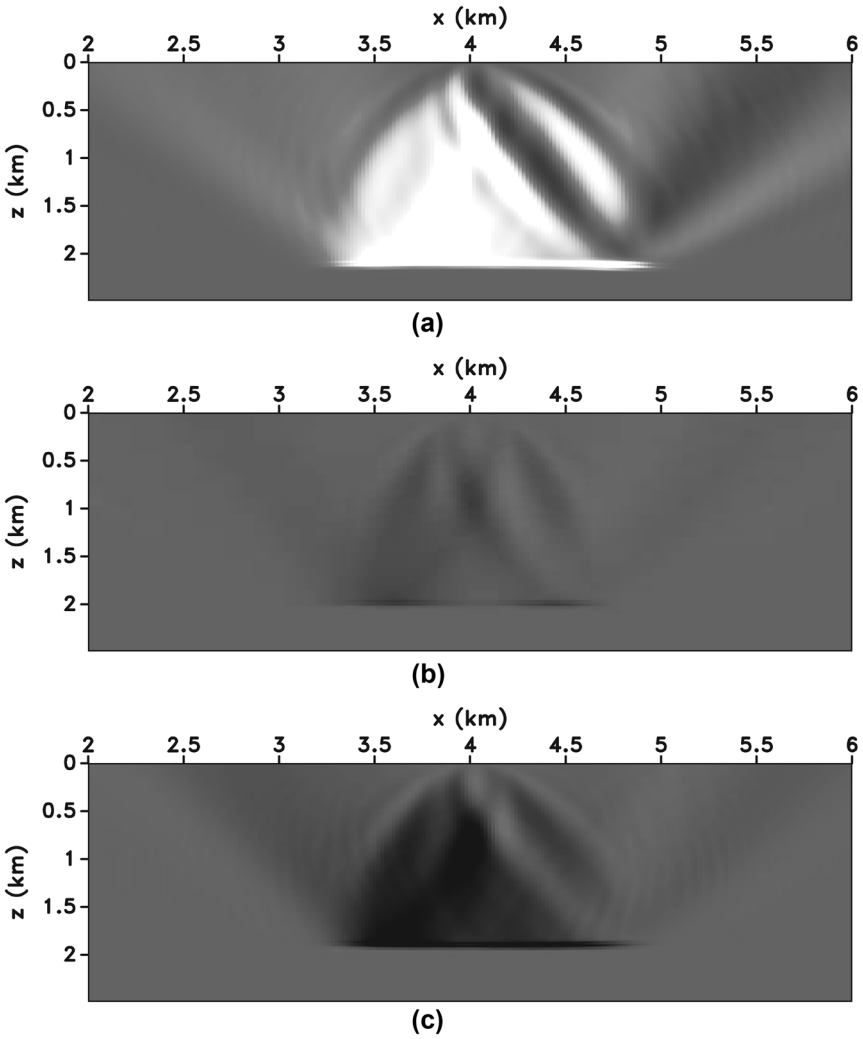


Figure 5 Gradient of the correlation objective function for a slowness model that is too low (a), correct (b) and too high (c). The slowness squared error is constant everywhere and equal to 10% of the background. The distance between the two shots is 60 m. We consider three shots and computed the gradient using the wavefields of the central experiment.

The shifts $r(\mathbf{x})$ in equation (4) are computed from the migrated images $R_i(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$. The dependency on the source and receiver wavefields and then the velocity model, is hidden in the imaging condition we use to construct the migrated image. An image of the subsurface $R(\mathbf{x})$ is conventionally computed as the zero-lag time correlation of the source and receiver wavefields $u_s(\mathbf{x}, t)$ and $u_r(\mathbf{x}, t)$ (Claerbout 1985):

$$R(\mathbf{x}) = \int_T u_s(\mathbf{x}, t)\, u_r(\mathbf{x}, t)\, dt, \tag{5}$$

where T is the recording time interval in the data.

We define the objective function:

$$\mathcal{J}(m) = \frac{1}{2} \sum_i \left\| \int P(\mathbf{x}, \boldsymbol{\lambda})\, c_i(\mathbf{x}, \boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right\|_{\mathbf{x}}^2, \tag{6}$$

where m represents the model parameters. Figure 3 shows the values of the objective function for different constant perturbations of the model used for the simple example shown in Fig. 1. The objective function is smooth and convex in the range of errors considered.

Computation of the gradient with the adjoint-state method

Because of the dimensionality of the model, the optimization problem is addressed using a local gradient-based method (steepest descent, conjugate gradient, etc.) (Vogel 2002). The computation of the derivative of the state variables with respect to the model parameters (Fréchet derivatives) is not practical (or even possible) because of the large number of dimensions of the model space. The adjoint-state method (Lions

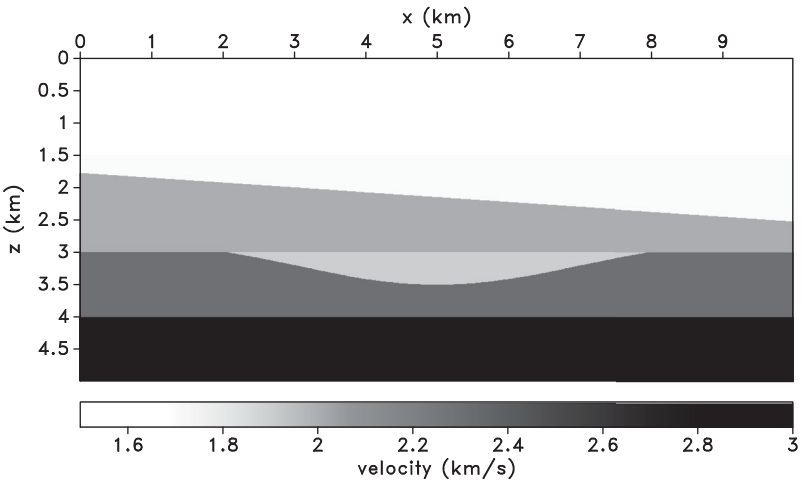


Figure 6 Slowness model used to generate the full-acoustic data. Absorbing boundary conditions have been applied at all sides of the model. The lighter shades of grey indicate faster layers. The top layer has slowness of water (0.66 s/km).

1972; Plessix 2006) is an efficient algorithm that computes the gradient of an objective function, which depends on some variables that describe the state of the physical system under analysis (state variables), without computing the Fréchet derivatives of these variables with respect to the model parameters.

The adjoint-state method consists in four steps: the computation of the state variables, the computation of the adjoint sources, the computation of the adjoint-state variables, and finally the computation of the gradient of the objective function. The first and third steps (the evaluation of the state and adjoint-state variables) require the solution of the wave-equation that governs the physics of the problem. The adjoint-state method is computationally efficient because only two wavefield simultations are required at each iteration for computing the gradient of the objective function instead of a simulation for each parameter in the model.

In image-domain wavefield tomography, one has two pairs of wavefields that are associated with the source and receiver side. For each experiment i, the source wavefield $u_{s,i}(\mathbf{x}, t)$ and receiver wavefield $u_{r,i}(\mathbf{x}, t)$ are obtained by solving the wave-equations:

$$\mathcal{L}(m)\, u_{s,i} = f_{s,i}, \tag{7}$$

$$\mathcal{L}^*(m)\, u_{r,i} = f_{r,i}, \tag{8}$$

where $f_{s,i}(\mathbf{x}, t)$ is the source signature and $f_{r,i}(\mathbf{x}, t)$ represents the reflection data for the ith experiment; $\mathcal{L}(m)$ and $\mathcal{L}^*(m)$ are the forward and adjoint modelling operator, respectively and m indicates the model parameters. Here, we consider a two-way wave operator (d'Alembert or Helmholtz) and m represents the squared slowness.

The source-side and receiver-side adjoint sources $g_{s,i}(\mathbf{x}, t)$ and $g_{r,i}(\mathbf{x}, t)$ are the partial derivatives of the objective function with respect to the associated wavefields:

$$g_{s,i} = \frac{\partial \mathcal{J}}{\partial u_{s,i}}, \tag{9}$$

$$g_{r,i} = \frac{\partial \mathcal{J}}{\partial u_{r,i}} \tag{10}$$

We solve the adjoint problem and compute the adjoint source and receiver wavefields $a_{s,i}(\mathbf{x}, t)$ and $a_{r,i}(\mathbf{x}, t)$ using the adjoint sources $g_{si}(\mathbf{x}, t)$ and $g_{r,i}(\mathbf{x}, t)$ as the force term:

$$\mathcal{L}^*(m)\, a_{s,i} = g_{s,i}, \tag{11}$$

$$\mathcal{L}(m)\, a_{r,i} = g_{r,i}. \tag{12}$$

Finally, the gradient is given by the inner product:

$$\frac{\partial \mathcal{J}}{\partial m} = -\left\langle a_{s,i}, \frac{\partial \mathcal{L}}{\partial m} u_{s,i} \right\rangle - \left\langle a_{r,i}, \frac{\partial \mathcal{L}^*}{\partial m} u_{r,i} \right\rangle. \tag{13}$$

For example, for the Helmholz operator (and its adjoint) parametrized in terms of slowness squared, $\mathcal{L} = -\nabla^2 - m\omega^2$, the partial derivative $\frac{\partial \mathcal{L}}{\partial m}$ returns a simple scaling factor $-\omega^2$ and equation (13) can also be seen as the sum of the scaled zero-lag correlation between $u_{s,i}$ and $a_{s,i}$ and between $u_{r,i}$ and $a_{r,i}$, similar to the procedure employed in conventional FWI but now involving both the source and receiver wavefields.

Adjoint sources

The particular design of the objective function impacts the adjoint sources and characterizes the particular wavefield

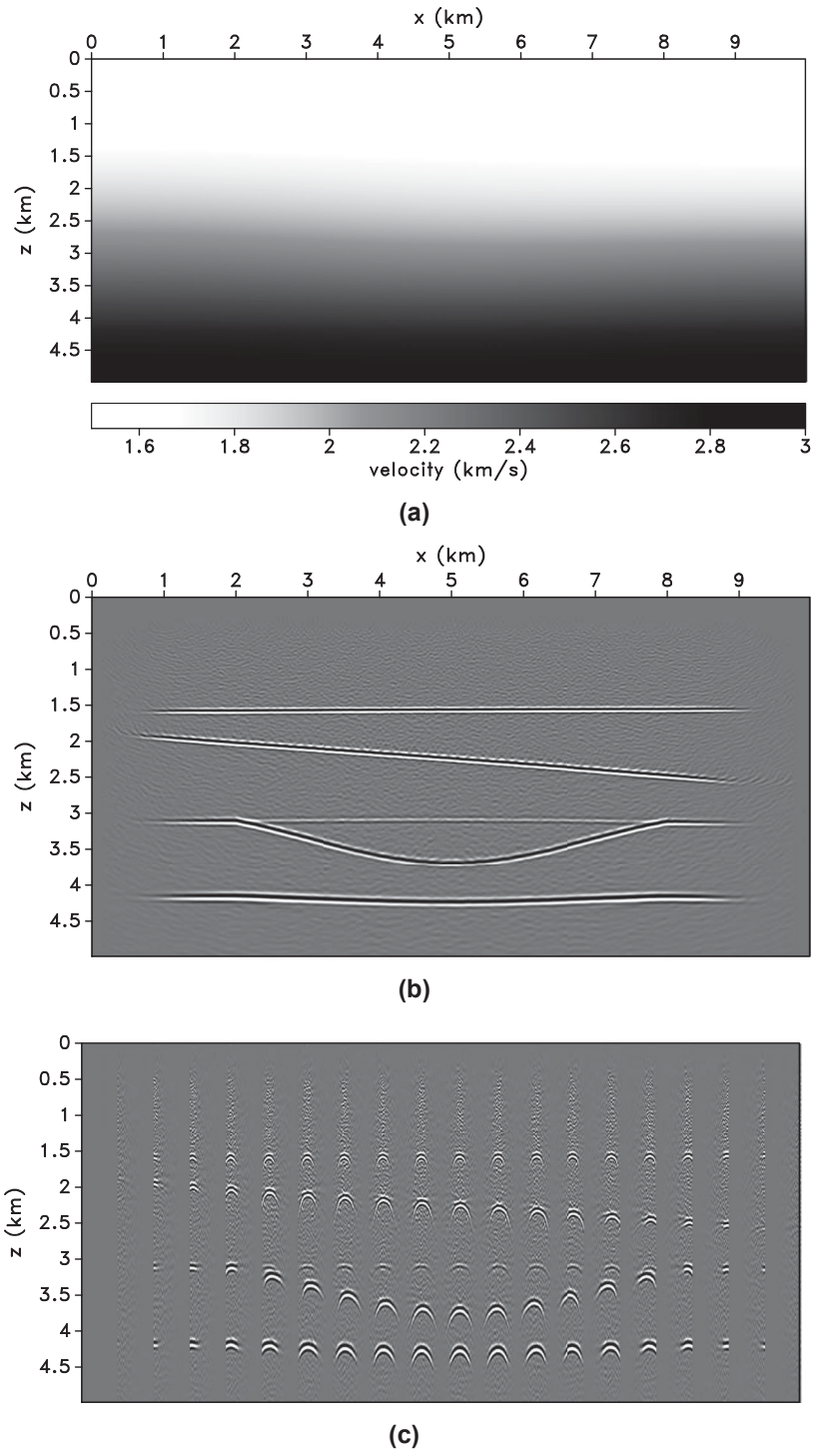


Figure 7 Initial slowness model (a), associated migrated image (b) and shot-domain common-image gathers. The reflectors are severely defocused and mispositioned. Observe the curvature of the common-image gathers that shows how different experiments image the reflectors at different depths.

tomography strategy proposed. In our objective function, we consider the dip field a slowly varying function of state variables and neglect the derivative of the penalty operator with respect to u. The details of the derivation of the expressions for the adjoint sources are in Appendix; here, we report the final result:

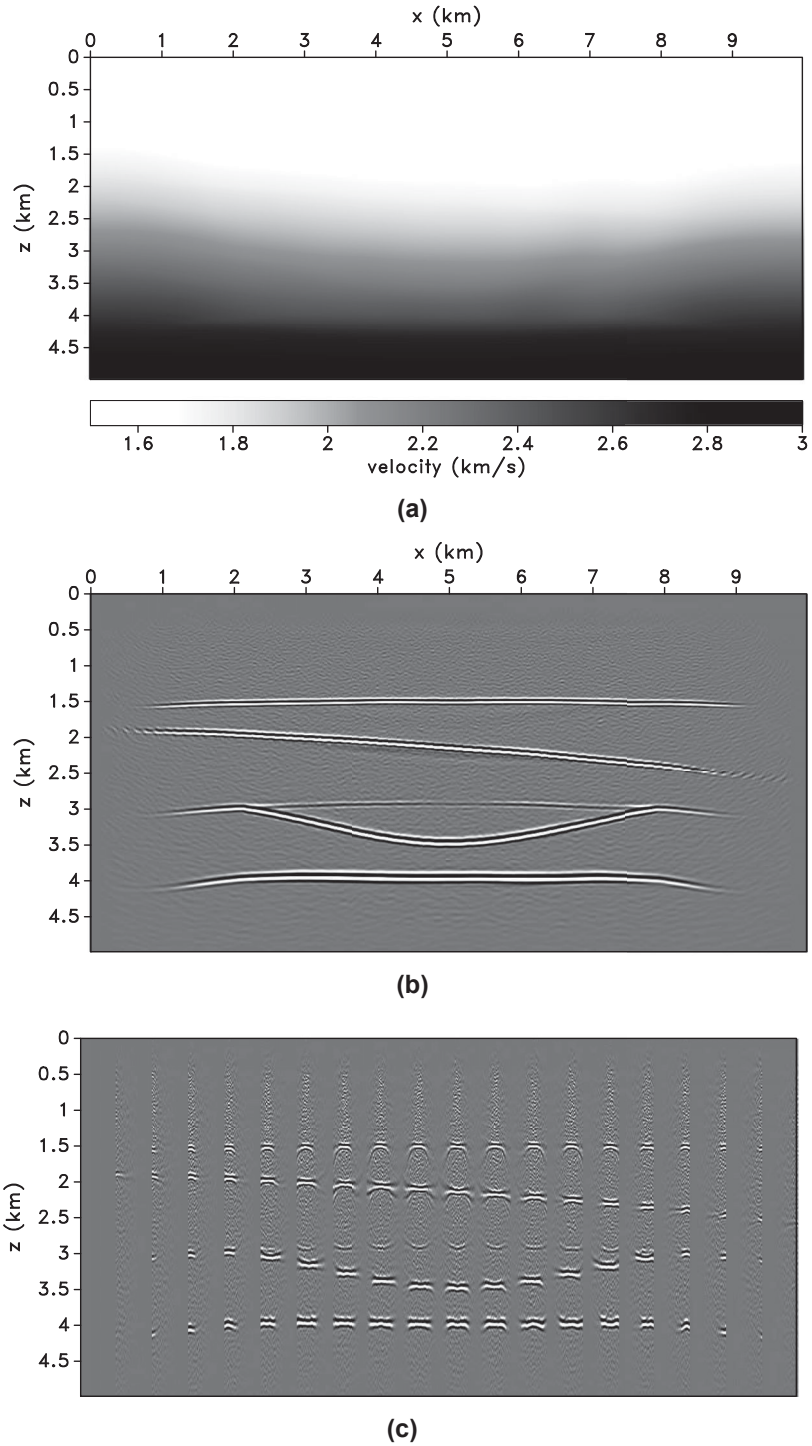


Figure 8 Slowness model after 15 iterations of waveform inversion (a), migrated image (b) and shot-domain common-image gathers (c). Observe the improved focusing and positioning of the reflectors. The common-image gathers show that the images are invariant in the shot direction and the model is thus kinematically correct. Observe the profile of the various layers in the reconstructed model.

$$g_{s,i} = u_{r,i}(\boldsymbol{\xi}, \mathrm{t}) \left[r_{i-1} \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) R_{i-1}(\boldsymbol{\xi} - \boldsymbol{\lambda})\ d\boldsymbol{\lambda} + r_i \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}(\boldsymbol{\xi} + \boldsymbol{\lambda})\ d\boldsymbol{\lambda} \right] \tag{14}$$

and

$$g_{r,i} = u_{s,i}(\boldsymbol{\xi}, \mathrm{t}) \left[r_{i-1} \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) R_{i-1}(\boldsymbol{\xi} - \boldsymbol{\lambda})\ d\boldsymbol{\lambda} + r_i \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}(\boldsymbol{\xi} + \boldsymbol{\lambda})\ d\boldsymbol{\lambda} \right], \tag{15}$$

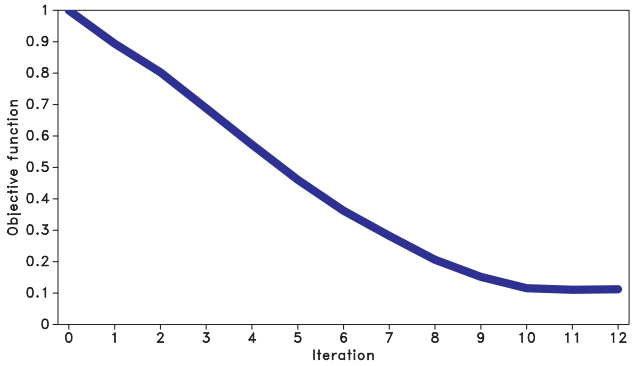


Figure 9 Evolution of the objective function with iterations. Observe the sharp decrease of the objective function and the rapid convergence to a kinematically accurate model.

where $g_{s,i} = g_{s,i}(\mathbf{x}, t)$ and $g_{r,i} = g_{r,i}(\mathbf{x}, t)$ indicate the adjoint sources for the source and receiver wavefield of the ith shot, respectively and $r_i = r_i(\mathbf{x})$ is defined as in equation (4).

In equations (14) and (15), the value of the residual $r(\mathbf{x})$ is spread over the local window spanned by the variable ξ around the image point $\boldsymbol{x}$ and weighted by the integral over λ, which represents a local convolution of the image and the penalty operator. Because of the linearity of the wave-equation with respect to the source term, we can then superimpose the contributions from different image points to construct the final adjoint sources. The value of $r(\mathbf{x})$ in equation (4) measures the relative displacement of one image with respect to the other; the adjoint source thus estimates the curvature of the moveout in the shot-domain common-image gathers at each image point and scales the background wavefields by this value.

In 3D, equations (14) and (15) remain valid but all the quantities involved are defined in a 3D space. The main difference between the two- and three-dimensional case is the computation of $r(\mathbf{x})$ and the ordering of the shots. In 2D there exists a clear ordering (along the line) for the seismic experiments, whereas in 3D every shot can have a number of neighbouring experiments.

The dependence of the penalty operator $P(\mathbf{x}, \boldsymbol{\lambda})$ on the state variables u passes through the definition of the dip vector $\boldsymbol{\nu}(\mathbf{x})$, which is normal to the reflectors at every point in the image. The dip field defines the wavefront set of the function that represents the reflection events in the image (Chang, Carrion and Beylkin 1987); the link between the image (and the wavefields) and the dip field can be written exploiting the concept of the gradient square tensor (van Vliet and Verbeek 1995): the dip vector is the eigenvector associated with the largest eigenvalues. There is no simple linear relationship between the dip vector and the wavefields, i.e., the state variables. Neglecting this term in the computation of the adjoint sources may introduce an error of the same order of magnitude of the term considered. A thorough study of $\partial P/\partial u$ and its impact on the computation of both the adjoint sources and the gradient of the objective function is subject for future research.

Flat reflector in a constant velocity medium

Figure 1 shows the migrated images of a horizontal interface obtained from a shot located at $x = 4$ km. The data are computed in a constant 2.0 km/s velocity medium using a two-way finite-difference scheme. We migrate the data with three different slowness models and perform sensitivity analysis for our methodology. Observe that with a single shot, the image of a horizontal reflector curves up or down depending on the sign of the velocity error. The relative displacement between different images is measured by the value of $r(\mathbf{x})$ defined in equation (4). Figure 4 shows the shifts calculated using two images from adjacent shots. Observe the change in sign across the zero-offset reflection point and how the sign changes with the velocity error. Note also the effect of limited aperture on the shift estimation: the abrupt change in the estimated shift at the edges of the reflector is due to uneven illumination from the three experiments.

In Fig. 5, we show the gradient computed from the correlation objective function. We consider three shots to compute $r(\mathbf{x})$ and use only the wavefields of the central experiment to compute the gradient. We also focus our attention in the portion of the reflector that is illuminated by all three shots in order to avoid artefacts due to the erroneous shift estimation at the edge of the reflector. The gradient of the objective function represents the direction of maximum increase of the function itself and its sign depends on the chosen parametrization. We compute the gradient with respect to slowness squared; this means that a positive gradient points toward slower models whereas a negative gradient points toward faster models.

EXAMPLES

We show the results of our inversion methodology using two data sets. First, we run an inversion test on a simple synthetic heterogeneous model with different layers using 40 shots evenly spaced 200 m apart and receivers at every grid

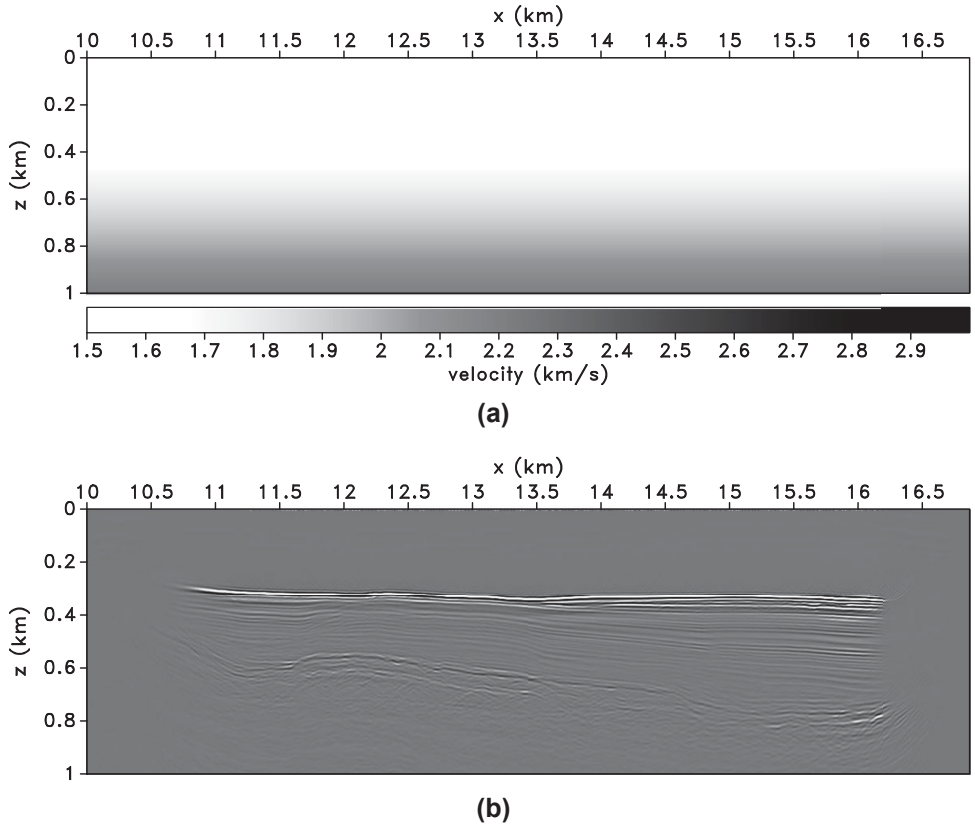


Figure 10 Initial migration model and migrated image.

point on the surface. We use shot-domain common-image gathers to assess the quality of the result and to show that our methodology is able to obtain a set of images that is invariant with respect to the experiment position. Second, we apply our methodology to a real 2D marine data set. We invert 200 shot-gathers acquired with a streamer cable. Each shot includes 99 receivers. The shot and receiver sampling is 12.5 m and the maximum offset is 1.225 km.

Synthetic laterally heterogeneous model

We use the synthetic heterogeneous model in Fig. 6 to test our inversion algorithm. The model has a few layers with different dips and a syncline structure with an inversion in the otherwise decreasing slowness trend. We generate full-acoustic (i.e., non-Born) data with absorbing boundary conditions (no free-surface multiples) and random Gaussian noise with a signal-to-noise ratio of 10. We model 40 shots evenly spaced 200 m apart with receivers at every grid point on the surface. We heavily smooth the initial model in Fig. 6 and obtain the initial slowness model for migration (Fig. 7a). The source signal is a Ricker wavelet with peak frequency 15 Hz. The modelled data are then band-pass filtered in order to limit the maximum frequency to 25 Hz. Figure 7(b) shows the stack of the migrated images obtained from the initial model; if we compare the correct model in Fig. 6 and the migrated image we can observe the mispositioning of the reflecting interfaces. We construct 10 shot-domain common image gathers to assess the overall focusing of the image. These gathers are simply the juxtaposition of the migrated images at fixed lateral positions (from $x = 0$ km to $x = 9$ km every 1 km). Figure 7(c) shows such gathers for the initial model: the position of the reflectors changes as a function of experiment (the images do not satisfy the semblance principle) and the model is thus not accurate. After 13 waveform tomography steps, we recover the model in Fig. 8(a). The model is still quite smooth but the migrated image shows a noticeable improvement in the position of the reflectors (Fig 8b), and the shot-domain common-image gathers in Fig. 8(c) clearly indicate better focusing. Moreover, the imaged reflectivity is now consistent with the actual slowness contrasts in the exact model; observe the change in polarity of the reflector at $z = 3$ km between the initial and final image,

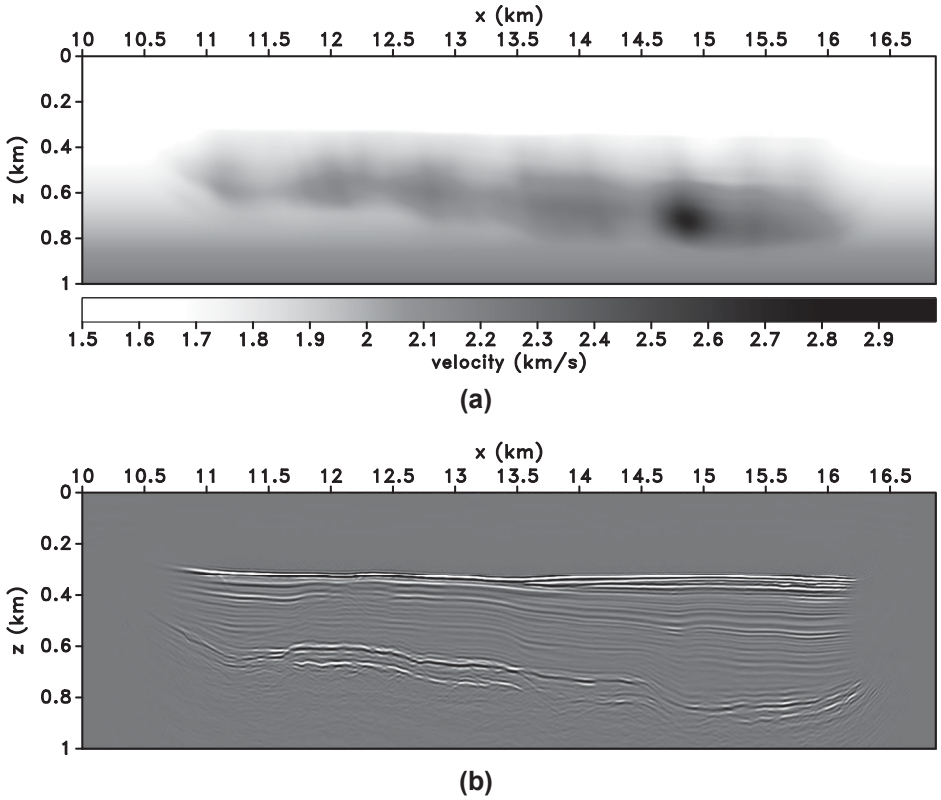


Figure 11 Final migration model and migrated image. The sections in Fig. 10(b) and Fig. 11(b) are displayed on the same colour scale. Observe the increased amplitude of the reflectors due to better focusing.

which corresponds to a decrease in velocity. We impose conformity between the model update and the layers in the image by means of a structure-oriented smoothing operator, which steers the gradient of the objective function by smoothing along the reflector slope. We used the complete stacked image to guide the smoothing of the gradient. This approach is conceptually similar to the sparse inversion proposed by Ma *et al.* (2010). Figure 9 shows the evolution of the objective function with iterations. Notice the monotonic decrease in the value of the residual and the flattening when the algorithm converges. The algorithm stops when the objective function changes less than 1% of the initial value over three consecutive iterations. The inversion shows a rapid convergence in the very first iterations of steepest-descent: the structure-oriented smoothing forces the gradient to conform to the geometry of the layers and speeds up convergence to an acceptable model.

Real data

We test our inversion algorithm on a real data set supplied by Eni E&P. The data consist in a single 2D line, the total number of shots is 3661, the streamer has 99 receivers, and the maximum offset is 1.225 km. The data have been regularized and demultipled. Source and receiver spacing is equal to 12.5 m. We show a subset of the entire data set; we downsample the shots 2:1 and analyse 200 shots. The first shot location is $x = 11.25$ km. The frequency content of the data ranges between 3–75 Hz.

We fix the water velocity to 1.51 km/s and start inversion from a slowness vertical gradient (Fig. 10a). The migrated image using the initial velocity model is shown in Fig. 10(b). In order to better understand the error in the velocity model, we show the shot-domain CIGs at 4 different locations (Fig. 12a-d). The CIGs are extracted every kilometre from $x = 12$ km. The strong moveout function of the shot position indicates error in the velocity model.

Observe the partial overlap between different events because of the moveout in the image-gathers. The interference between moveout curves that is relative to different reflection events would cause a naive MVA approach, as the total stack power maximization, to cycle skip and not converge.

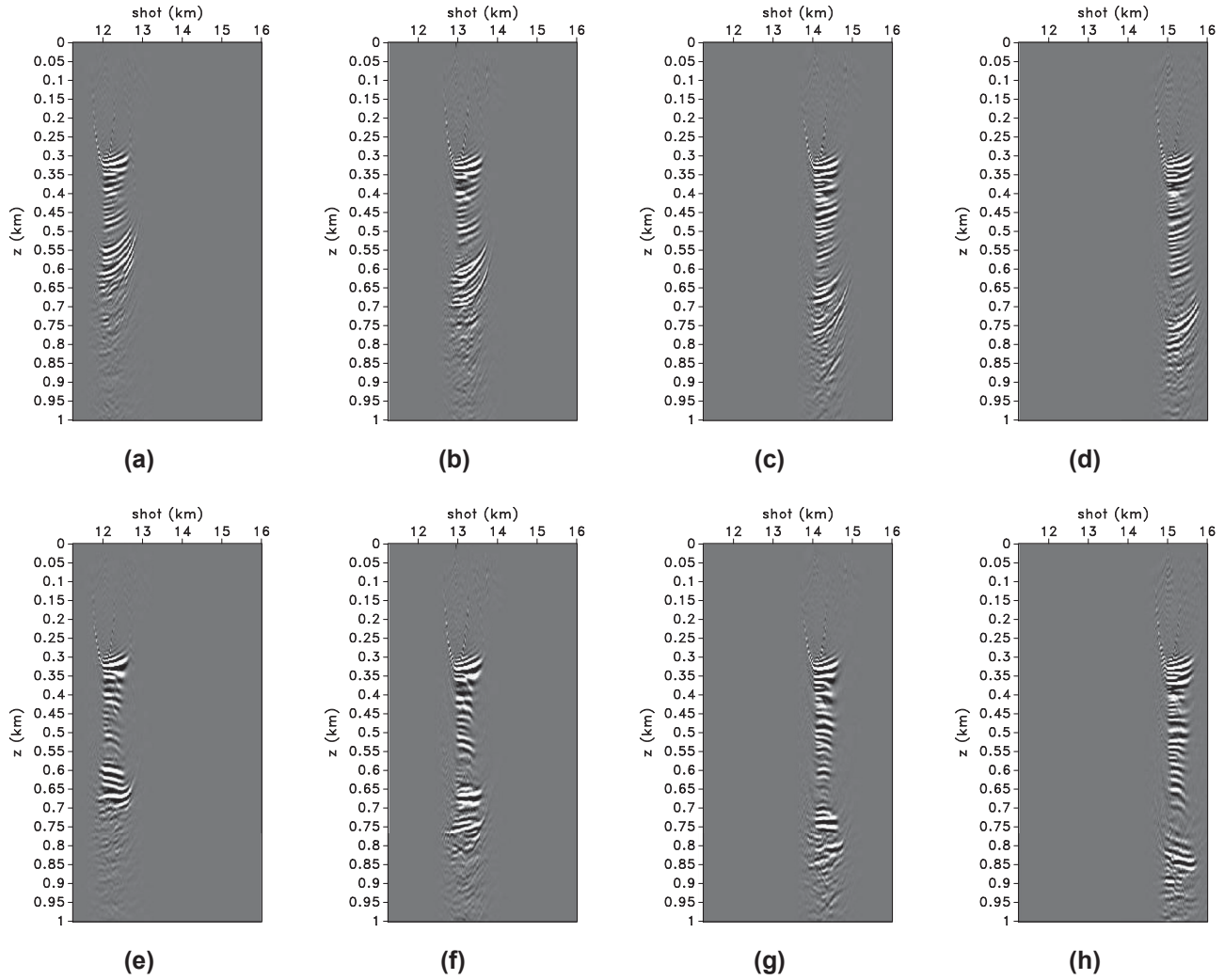


Figure 12 Shot-domain CIGs computed for the incorrect model at (a) $x = 12$ km, (b) $x = 13$ km, (c) $x = 14$ km and (d) $x = 15$ km and gathers computed after inversion at the same locations ((e), (f), (g) and (h)). The gathers are flattened by the inversion procedure, i.e., the reflectors are invariant with respect to the shot index.

The inversion converges after 90 iterations of steepest descent and returns the model in Fig. 11(a); the associated migrated image is shown in Fig. 11(b). The migrated images reconstructed with the initial and final models are displayed using the same scale. Observe the increased strength of the reflectors after inversion. The common-image gathers in Fig. 12(e-h) show that the single migrated images locate reflectors at consistent depths (in agreement with the semblance principle) and thus the velocity model is kinematically accurate.

We compute angle domain common-image gathers (ADCIGs) associated with the initial and final models using the method developed by Sava and Fomel (2003). The gathers are computed for opening angles between −25 and 0 degrees. Figure 13(a) shows the ADCIGs for the initial model: the strong moveout indicates errors in the velocity model. Figure 13(b) shows the ADCIGs after inversion. The events are flatter and indicate a kinematically more accurate model.

The convergence of the inversion algorithm is described by the evolution of the objective function (Fig. 14). No model regularization is implemented in the inversion. The gradient is smoothed along the reflectors using a structural-oriented smoothing. The width of the smoothing filters is 300 m in the direction perpendicular to the reflectors and 100 m i the direction across. Notice that because of the greater number of reflection events and the irregularity of the interfaces, many

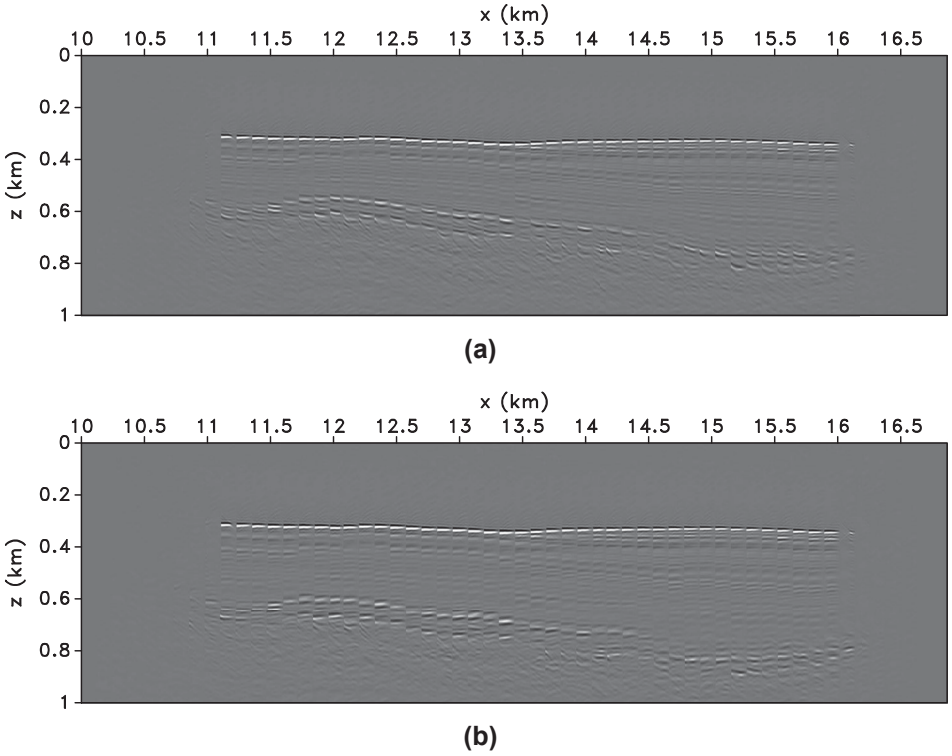


Figure 13 Angle-domain common-image gathers (ADCIGs) computed at several lateral positions (a) before and (b) after inversion. After inversion the gathers are flatter, which indicates a more kinematically accurate migration model.

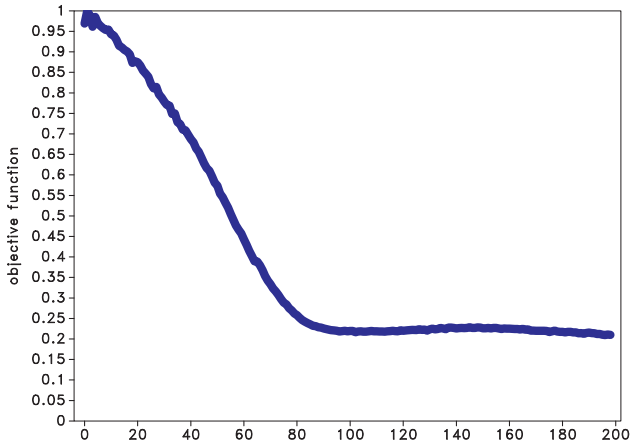


Figure 14 The evolution of the objective function shows that the algorithm converges after about 80 iterations.

more iterations are needed to reach convergence with respect to the previous synthetic example.

DISCUSSION

A complete image of the subsurface is the superposition of partial images from individual experiments. The semblance principle (Al-Yahya 1989) is a common criterion for assessing the correctness of the velocity model used for imaging the survey: when the velocity model is correct, all shots locate reflectors at the same position, i.e., the image is invariant along the experiment axis. Several shots are needed to evaluate a velocity error at a single point in space.

We can evaluate the invariance along the experiment axis by computing the energy of the first derivative in that dimension. The first derivative acts as a penalty operator by highlighting and enhancing deviations from the horizontal direction along the shot axis. The stack of the energy of the first derivative is the differential semblance operator applied directly in the shot domain (Symes 1991; Plessix 2006).

Here, we explore an alternative statement of the semblance principle: when the velocity model is correct, images from different shots constructively interfere and build up the image perpendicular to the structural dip or parallel to the reflector slope, at every point in the image space. The structural dip is a commonly extracted attribute and can be linked to the image itself by means of the gradient square tensors (van Vliet and Verbeek 1995). Unfortunately, the relationship

between wavefields and the dip field is non-linear: the dip vector represents the eigenvector associated with the largest eigenvalue of the gradient square tensor. The dip variation with respect to model perturbation in the inversion procedure is not straightforward and further research is needed to develop an efficient method to exploit this information for velocity analysis purposes. Nonetheless, we are able to measure the semblance of two images through appropriately penalized local correlations of pairs of images. If the velocity model is correct, the maximum of the local correlation is along the reflector at every point in the image; if the model is incorrect, the maximum deviates from the reflector slope. The penalty operator is space-dependent and annihilates the correlation panels orthogonal to the reflector dip. Because of the dependency on the velocity model, we measure the dip field at each tomographic iteration; the estimation can be carried out efficiently using gradient square tensors (van Vliet and Verbeek 1995).

The correlation objective function is superior to the image difference in many respects. First, the difference between two images depends on image amplitudes that change as a function of the shot position and cannot be matched point-wise as in the standard implementation of full-waveform inversion in the data space. The amplitude patterns affect the residual in the image space and effectively contribute to the adjoint source calculation, even if the velocity model is correct and the gradient of the objective function is zero (Mulder and Kroode 2002). By penalizing local correlations, we reduce the dependence of the objective function on amplitudes, thus increasing the robustness and reducing the systematic bias caused by the amplitude differences between the images. A downside of the correlation operator is the loss of spatial resolution. Correlation in the spatial domain is equivalent to multiplication in the dual frequency/wavenumber domain; for signals with both finite spatial support and bandwidth, the multiplication of the spectra decreases the bandwidth, i.e., increases the width of the signal in the spatial domain. To increase the resolution and accuracy of the evaluation of the relative shifts between images, deconvolution is a viable improvement over correlation. If we assume that two images from nearby experiments are linked by a simple spatial shift, a local deconvolution approach would ideally produce a band-limited spatial delta function indicating the direction of apparent displacement. On the other hand, since deconvolution amplifies the noise in the data (because of the division in the frequency domain), additional care is necessary to stabilize the result.

A thorough study of the effect of the shot distance on the computed gradient would be beneficial in relation to phenomena such as cycle skipping, which hinders many velocity analysis strategies (especially but not exclusively in the data domain). The image domain is intrinsically less prone to cycle skipping problems; nevertheless the question remains about what happens when two images illuminate rather different areas of the subsurface and the local correlations cannot be used to estimate a reliable and meaningful shift between the two images. From our current knowledge and understanding, the shots must be close enough to provide comparable images of the subsurface and avoid cycle skipping in the image domain. The sampling of the shot position and the illumination pattern in the subsurface can also create a scenario in which the two images do not overlap in certain local windows.

Our method is based on locally coherent events, such as locally smooth reflectors. Conflicting dips, fault planes and areas where the definition of a reflection plane is ambiguous represent open problems because in these areas we cannot define penalty operators. In this respect, plane-wave migration may represent a valid solution because of the implicit spatial filtering of the image: each plane-wave reconstructs a particular subset of the dips in the model, thus reducing the ambiguity in defining the penalty operators. Alternatively, for complex areas with conflicting dips and complicated geologic features (where there is no clear definition of the dip field), a more sophisticated design of the penalty operator may be key to obtaining meaningful residuals and a reliable gradient. The eigenvalues and eigenvectors of the gradient square tensors (van Vliet and Verbeek 1995) can be used to define ellipses that, in turn, may offer a structure-oriented criterion for the definition of the penalty operators.

In contrast with other techniques in the image space (Bishop *et al.* 1985; Biondi and Symes 2004; Lambaré *et al.* 2004; Xie and Yang 2008; Yang and Sava 2010 2011), our approach does not need to explicitly construct and analyse image gathers (or extended images). A further cost-saving factor is given by the selection of points on the imaged reflectors (Yang and Sava 2010) and away from complex areas (pinchouts or areas with conflicting dips). The computation of the local correlation is carried out at every image point using the efficient method developed by Hale (2006).

The ability to extract velocity information from pairs of experiments adds a new degree of freedom for implementing a model update procedure. Here, we indicate two possible strategies. We can simultaneously include all the shots in the survey in the definition of the objective function or we can proceed iteratively and update the velocity model using the information obtained from a single pair of experiments before moving to the next pair. Nearby shots probe similar portions

of the model and provide comparable images; we can use the information extracted from an initial group to update the model used for imaging a second group of experiments. Since in any migration velocity analysis scheme we have to image the entire survey at least once, the iterative update of the model over shots becomes cost-effective if we actually reduce the number of global migrations of the entire data set.

In 2D, we can easily define an order for the experiments; in the general 3D scenario, we have an extra degree of freedom. The definition of the objective function remains the same but a given shot can have more than 2 neighbouring experiments (the acquisition is defined on a two-dimensional grid). We can generalize the concept of correlation using the semblance functional (Taner and Koehler 1969), which is non-linear in the input signals and makes the computation of the gradient more involved. We can also separately analyse each pair of experiments given a reference model. Further research and numerical tests are needed to assess efficiency and robustness of one strategy over the other.

CONCLUSION

We develop an objective function for migration velocity analysis in the shot-image domain. Our methodology is based on local image correlations and a reformulation of the semblance principle involving only small groups of migrated images. The optimization criterion is the minimization of the apparent shift between the migrated images and naturally leads to a differential semblance optimization problem, which is iteratively solved using a gradient-based method. The linearity of the operators in the objective function makes the computation of the gradient practical using the adjoint-state method. Our approach is full-wave because we do not rely on the linearization of the wave extrapolation procedure with respect to the model parameters to construct a migration velocity analysis operator. We iteratively solve a non-linear problem with a gradient-based algorithm and the simulation of the wavefields is fully non-linear with respect to the model parameters.

The structural features in the migrated image and the variation of these features as a function of the experiment supply valuable information about the model parameters. Here, we showed that it is possible, without constructing common-image gathers , to extract information about the model from a very limited number of experiments using a warping relationship between migrated images. We used a few synthetic tests to describe the algorithm and the behaviour of the inversion procedure; we inverted a real 2D marine data set to demonstrate the effectiveness of our method.

ACKNOWLEDGEMENTS

We would like to thank Eni E&P for permission to publish this work and Nicola Bienati for the valuable conversations on tomography and inversion. The reproducible numerical examples in this paper use the Madagascar open-source software package freely available from http://www.reproducibility.org. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. Francesco Perrone would also like to thank Thomas Cullison for developing the automatic picking algorithm used in this research, Jeff Godwin for developing the interfaces for the cluster and Prof. Luis Tenorio for his help and mathematical support. We thank Dr. Yu Zhang and Dr. René-Edouard Plessix for their helpful and constructive comments that greatly improved the paper.

REFERENCES

Al-Yahya K. 1989. Velocity analysis by iterative profile migration. *Geophysics* **54**, 718–729.

Albertin U., Sava P., Etgen J. and Maharramov M. 2006. Adjoint wave-equation velocity analysis. Presented at the 74th Annual International Mtgeeting, SEG.

Baysal E., Kosloff D.D. and Sherwood J.W.C. 1983. Reverse time migration. *Geophysics* **48**, 1514–1524.

Biondi B. and Sava P. 1999. Wave-equation migration velocity analysis. 69th Annual International Meeting, SEG, Expanded Abstracts, 1723–1726.

Biondi B. and Symes W.W. 2004. Angle-domain common-image gathers for migration velocity analysis by wavefield continuation imaging. *Geophysics* **69**, 1283–1298.

Bishop T.N., Bube K.P., Cutler R.T., Langan R.T., Love P.L., Resnick J.R., et al. 1985. Tomographic determination of velocity and depth in laterally varying media. *Geophysics* **50**, 903–923.

Brossier R., Operto S. and Virieux J. 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics* **75**, R37–R46.

Carcione J. 2007. *Wave Fields in Real Media Wave Propagation in Anisotropic, Anelastic, Porous and Electromagnetic Media*. Elsevier. Handbook of Geophysical Exploration.

Chang W., Carrion P. and Beylkin G. 1987. Wavefront sets of solutions to linearised inverse scattering problems. *Inverse Problems* **3**, 683–690.

Chavent G. and Jacewitz C.A. 1995. Determination of background velocities by multiple migration fitting. *Geophysics* **60**, 476–490.

Claerbout J.F. 1985. *Imaging the earth's interior*. Blackwell Publishing.

Farmer P.A., Jones I.F., Zhou H., Bloor R.I. and Goodwin M.C. 2006. Application of reverse time migration to complex imaging problems. *First Break* **24**, 65–73.

Faye J.-P. and Jeannot J.-P. 1986. Prestack migration velocity analysis from focusing depth imaging. Presented at the 56th Annual International Meeting, SEG.

Fomel S. 2002. Applications of plane-wave destruction filters. *Geophysics* **67**, 1946–1960.

Fowler P. 1985. Migration velocity analysis by optimization. Technical Report 44, Stanford Exploration Project.

Gazdag J. and Sguazzero P. 1984. Migration of seismic data by phase shift plus interpolation. *Geophysics* **49**.

Hale D. 2006. Fast local cross-correlations of images. 76th Annual International Meeting, SEG, Expanded Abstracts, 3160–3164.

Hale D., Cox B. and Hatchell P.J. 2008. Apparent horizontal displacements in time-lapse seismic images. 78th Annual International Meeting, SEG, Expanded Abstracts, 3169–3173.

Kelly S., Martinez J.R., Tsimelzon B. and Crawley S. 2010. A comparison of inversion results for two full-waveform methods that utilize the lowest frequencies in dual-sensor recordings. Presented at the 80th Annual International Meeting, SEG.

Lambaré G., Alerini M., Baina R. and Podvin P. 2004. Stereotomography: a semi-automatic approach for velocity macromodel estimation. *Geophysical Prospecting* **52**, 671–681.

Lions J. 1972, *Nonhomogeneous Boundary Value Problems and Applications*. Springer Verlag, Berlin.

Liu F., Hanson D.W., Whitmore N.D., Day R.S. and Stolt R.H. 2006. Toward a unified analysis for source plane wave migration. *Geophysics* **71**, S129–S139.

Ma Y., Hale D., Meng Z. and Gong B. 2010. Full waveform inversion with image-guided gradient. Presented at the 80th Annual International Meeting, SEG.

McMechan G.A. 1983. Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting* **31**, 413–420.

Morton S.A. and Ober C.C. 1998. Faster shot-record migration using phase encoding. 68th Annual International Meeting, SEG, Expanded Abstracts, 1131–1134.

Mulder W.A. and Kroode A.P.E. 2002. Automatic velocity analysis by differential semblance optimization. *Geophysics* **67**, 1184–1191.

Perrone F. and Sava P. 2012a. Wave-equation migration with dithered plane waves. *Geophysical Prospecting* **60**, 444–465.

Perronem F. and Savam P. 2012b. Waveform tomography based on local image correlations. 82th Annual International Meeting, SEG, Expanded Abstracts.

Plessix R.-E. 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International* **167**, 495–503.

Pratt R.G. 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics* **64**, 888–901.

Rickett J. and Sava P. 2002. Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics* **67**, 883–889.

Romero L.A., Ghiglia D.C., Ober C.C. and Morton S.A. 2000. Phase encoding of shot records in prestack migration. *Geophysics* **65**, 426–436.

Santosa F. and Symes W.W. 1989. An analysis of least-squares velocity inversion. Geophysical Monograph no. 4. Society of Exploration Geophysicists.

Sattlegger J.W. 1975. Migration velocity determination: Part I. Philosophy. *Geophysics* **40**, 1–5.

Sava P., Biondi B. and Etgen J. 2005. Wave-equation migration velocity analysis by focusing diffractions and reflections. *Geophysics* **70**, U19–U27.

Sava P. and Fomel S. 2003. Angle-domain common-image gathers by wavefield continuation methods. *Geophysics* **68**, 1065–1074.

Sava P. and Vasconcelos I. 2009. Extended common-image-point gathers for wave-equation migration. Presented at the 71st Conference and Exhibition, EAGE.

Sirgue L., Barkved O.I., Dellinger J., Etgen J., Albertin U. and Kommedal J.H. 2010. Full waveform inversion: the next leap forward in imaging at Valhall. *First Break* **28**, 65–70.

Sirgue L. and Pratt R. 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics* **69**, 231–248.

Soubaras R. 2006. Modulated-shot migration. 76th Annual International Meeting, SEG, Expanded Abstracts, 2430–2434.

Stoffa P.L., Fokkema J.T., de Luna Freire R.M. and Kessiger W.P. 1990. Split-step Fourier migration. *Geophysics* **55**, 410–421.

Stoffa P.L., Sen M.K., Seifoullaev R.K., Pestana R.C. and Fokkema J.T. 2006. Plane-wave depth migration. *Geophysics* **71**, S261–S272.

Stolk C.C. and Symes W.W. 2004. Kinematic artifacts in prestack depth migration. *Geophysics* **69**, 562–575.

Stolt R.H. 1978. Migration by Fourier transform. *Geophysics* **43**, 23–48.

Symes W.W. 1991. Layered velocity inversion: A model problem from reflection seismology problem from reflection seismology. *SIAM Journal on Mathematical Analysis* **22**, 680–716.

Symes W.W. 1993. A differential semblance criterion for the inversion of multioffset seismic reflection data. *Journal of Geophysical Research* **98**, 2061–2073.

Symes W.W. 2008. Migration velocity analysis and waveform inversion. *Geophysical Prospecting* **56**, 765–790.

Symes W.W. and Carazzone J.J. 1991. Velocity inversion by differential semblance optimization. *Geophysics* **56**, 654–663.

Taner M.T. and Koehler F. 1969. Velocity spectra - digital computer derivation and applications of velocity functions. *Geophysics* **34**, 859–881.

Tarantola A. 1984. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **49**, 71–92.

Thomsen L. 1986. Weak elastic anisotropy. *Geophysics* **51**, 1954–1966.

van Vliet L.J. and Verbeek P.W. 1995. Estimators for orientation and anisotropy in digitalized images. *Proceedings of the first annual conference of the Advanced School for Computing and Imaging, ASCI'95*, 442–450.

Vogel C. 2002. *Computational methods for inverse problems*. SIAM.

de Vries D. and Berkhout A.J. 1984. Velocity analysis based on minimum entropy. *Geophysics* **49**, 2132–2142.

Whitmore N. 1995. An Imaging Hierarchy for Common-Angle Plane Wave Seismograms. PhD thesis, University of Tulsa.

Woodward M.J. 1992. Wave-equation tomography. *Geophysics* **57**, 15–26.

Xie X.B. and Yang H. 2008. The finite-frequency sensitivity kernel for migration residual moveout and its applications in migration velocity analysis. *Geophysics* **73**, S241–S249.

Yang T. and Sava P. 2010. Wave-equation migration velocity analysis with extended common-image-point gathers. 80th Annual International Meeting, SEG, Expanded Abstracts, 4369–4374.

Yang T. and Sava P. 2011. Wave-equation migration velocity analysis with time-shift imaging. *Geophysical Prospecting* **59**, 635–650.

Yilmaz O. and Chambers R. 1984. Migration velocity analysis by wave-field extrapolation. *Geophysics* **49**, 1664–1674.

Zhang Y., Sun J., Notfors C., Gray S.H., Chernis L. and Young J. 2005. Delayed-shot 3d depth migration. *Geophysics* **70**, E21–E28.

Zhang Y., Xu S., Tang B., Bai B., Huang Y. and Huang T. 2010. Angle gathers from reverse time migration. *The Leading Edge* **29**(11), 1364–1371.

APPENDIX

COMPUTATION OF THE SOURCE AND RECEIVER ADJOINT SOURCE USING PERTURBATION THEORY

Our objective function is defined as:

$$\mathcal{J}(m) = \frac{1}{2}\sum_i \|r_i(\mathbf{x})\|_{\mathbf{x}}^2$$

$$= \frac{1}{2}\sum_i \left\| \int P(\mathbf{x},\boldsymbol{\lambda})\, c_i(\mathbf{x},\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right\|_{\mathbf{x}}^2, \tag{A1}$$

where $c_i(\mathbf{x},\boldsymbol{\lambda}) = \int w(\mathbf{x}-\boldsymbol{\xi}) R_{i+1}(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2}) R_i(\boldsymbol{\xi}+\frac{\boldsymbol{\lambda}}{2}) d\boldsymbol{\xi}$ is the local correlation of two images from the nearby shots i and $(i+1)$ and $P(\mathbf{x},\boldsymbol{\lambda})$ is an operator that annihilates the local correlation panel along the direction of the structure at every point in the image.

The adjoint-state method computes the gradient of the objective function with respect to the model parameters by solving two forward problems for each state variable of the system. This procedure turns out to be extremely computationally efficient since it avoids perturbing each model parameter and computing the resulting state variable perturbation (Lions 1972; Plessix 2006). The solution of the auxiliary forward problem requires different sources (adjoint sources) that are computed by differentiating the objective function with respect to the state variables.

We now derive the mathematical expression for the source and receiver adjoint sources. The computation of the derivative of a functional with respect to a function requires the definition of an inner product in the appropriate functional space; moreover, it requires the concept of the Fréchetdifferential (Vogel 2002).

Let us denote the state variables of the problem, i.e., the source and receiver wavefields for the ith shot, by $u_{s,i}$ and $u_{r,i}$, respectively. We derive the expression of the adjoint source for the source wavefield. The derivation and expression of the adjoint source for the receiver wavefield are analogous. First, we make explicit the dependency of the residual $r_i(\mathbf{x})$ (and thus of the objective function $\mathcal{J}$) on the state variable:

$$r_i(\mathbf{x}) = \int P(\mathbf{x},\boldsymbol{\lambda})\, w(\mathbf{x}-\boldsymbol{\xi})\, u_{s,i}\left(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2},\mathrm{t}\right)\cdot$$

$$\cdot\; u_{r,i}\left(\boldsymbol{\xi}-\frac{\boldsymbol{\lambda}}{2},\mathrm{t}\right) R_{i+1}\left(\boldsymbol{\xi}+\frac{\boldsymbol{\lambda}}{2}\right) dV, \tag{A2}$$

where the integral is a triple integral over the volume element $dV = dt d\xi \mathbf{d}\lambda$. Second, we compute the Fréchetdifferential of the objective function $\mathcal{J}(m)$ with respect to an arbitrary perturbation of the state variable $\delta u_{s,i}$. The objective function in equation (A1) is the sum over shots of energy of the residues in equation (A2) over the spatial domain. In order to simplify the derivation, we first write the differential of the objective function as a function of the differential of the residual using perturbation theory:

$$\mathcal{J}[u_{s,i}] + \delta\mathcal{J}[u_{s,i},\delta u_{s,i}] = \frac{1}{2}\int \left(r_{i-1} + \delta r_{i-1}[u_{s,i},\delta u_{s,i}]\right)^2$$

$$+ \left(r_i + \delta r_i[u_{s,i},\delta u_{s,i}]\right)^2 d\mathbf{x}. \tag{A3}$$

Expanding the squares in equation (A3) and considering only the terms that are linear in the wavefield perturbation, we obtain an expression for the Fréchetdifferential of the objective function:

$$\delta\mathcal{J}[u_{s,i},\delta u_{s,i}] = \int r_{i-1}\delta r_{i-1}[u_{s,i},\delta u_{s,i}]$$

$$+ r_i \delta r_i[u_{s,i},\delta u_{s,i}] d\mathbf{x}. \tag{A4}$$

The square brackets in equation (A4) indicate the double dependency of the Fréchetdifferential on the state variable and its arbitrary perturbation. The Fréchetdifferential can be expressed as the inner product between the Fréchetderivative $\nabla_{u_{s,i}}\mathcal{J}$ of the objective function with respect to the state variable $u_{s,i}$ and the arbitrary perturbation $\delta u_{s,i}$:

$$\delta\mathcal{J} = \langle \nabla_{u_{s,i}}\mathcal{J}, \delta u_{s,i}\rangle. \tag{A5}$$

Our goal is then to express the Fréchetdifferential as an operator applied to the arbitrary perturbation of the state variable; that operator is the Fréchetderivative.

Let us derive the expression of the Fréchetdifferential of the residual with respect to the state variable $\delta r_i[u_{s,i},\delta u_{s,i}]$ and

plug it into the expression of the Fréchetdifferential of the objective function. Assuming $r_i(\mathbf{x})$ is Fréchetdifferentiable, we compute the Fréchetdifferential using the limit of the incremental ratio:

$$\delta r_i[u_{s,i}, \delta u_{s,i}] = \lim_{h\to 0} \frac{r_i[u_{s,i} + h\delta u_{s,i}] - r_i[u_{s,i}]}{h}, \tag{A6}$$

where h is a scalar and $\delta u_{s,i}$ is an arbitrary perturbation of the state variable $u_{s,i}$ and belongs to the same functional space. The residual r_i is a linear functional of $u_{s,i}$, and computing the limit in equation (A6) we obtain:

$$\delta r_i[u_{s,i}, \delta u_{s,i}] = \int P(\mathbf{x}, \boldsymbol{\lambda})\, w(\mathbf{x} - \boldsymbol{\xi})\, \delta u_{s,i}\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}, t\right) \cdot$$

$$\cdot\ u_{r,i}\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}, t\right) R_{i+1}\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) dV. \tag{A7}$$

The expression for $\delta r_{i-1}[u_{s,i}, \delta u_{s,i}]$ is analogous and can be derived using the same procedure.

We want to express equation (A7) as an inner product between an operator (the Fréchetderivative $\nabla_{u_{s,i}} r_i$) and the state variable perturbation $\delta u_{s,i}$; by exploiting the sifting property of the Dirac delta function, we can rewrite equation (A7) in order to have the wavefield perturbation $\delta u_{s,i}$ depend on only one spatial and one temporal variable:

$$\delta r_i[u_{s,i}, \delta u_{s,i}] = \int P(\mathbf{x}, \boldsymbol{\lambda})\, w(\mathbf{x} - \boldsymbol{\xi})\, \delta\left(\boldsymbol{\eta} - \left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right)\right) \delta u_{s,i}(\boldsymbol{\eta}, \mathrm{t}) \cdot$$

$$\cdot\ u_{r,i}(\boldsymbol{\eta}, t)\, R_{i+1}\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) dV d\boldsymbol{\eta}. \tag{A8}$$

We can express equation (A8) as an inner product in the form:

$$\delta r_i = \left\langle \nabla_{u_{s,i}} r_i, \delta u_{s,i} \right\rangle, \tag{A9}$$

where the inner product is over the domain of the perturbation (η, t). From equations (A8) and (A9), the expression of $\nabla_{u_{s,i}} r_i$ follows:

$$\nabla_{u_{s,i}} r_i = u_{r,i}(\boldsymbol{\eta}, \mathrm{t}) \int P(\mathbf{x}, \boldsymbol{\lambda})\, w\left(\mathbf{x} - \boldsymbol{\eta} - \frac{\boldsymbol{\lambda}}{2}\right) R_j(\boldsymbol{\eta} + \boldsymbol{\lambda})\, d\boldsymbol{\lambda}, \tag{A10}$$

where we exploited again the sifting property of the Dirac delta function and integrated over $d\xi$.

We can now substitute the expression for δr_i and δr_{i-1} in the equation for $\delta \mathcal{J}$:

$$\delta \mathcal{J}[u_{s,i}, \delta u_{s,i}] = \int r_{i-1}(\mathbf{x})\, \nabla_{u_{s,i}} r_{i-1}\, \delta u_{s,i}[u_{s,i}, \delta u_{s,i}]\, d\mathbf{x} d\boldsymbol{\eta} d\mathrm{t}$$

$$+ \int r_i(\mathbf{x})\, \nabla_{u_{s,i}} r_i\, \delta u_{s,i}[u_{s,i}, \delta u_{s,i}]\, d\mathbf{x} d\boldsymbol{\eta} d\mathrm{t}. \tag{A11}$$

By combining equations (A11) (A10) and (A5), we can isolate the expression of the Fréchetderivative of the objective function $\mathcal{J}$ with respect to the state variable $u_{s,i}$:

$$\nabla_{u_{s,i}} \mathcal{J} = u_{r,i}(\boldsymbol{\eta}, \mathrm{t})\, [r_{i-1} P(\mathbf{x}, \boldsymbol{\lambda}) * R_{i-1}(\mathbf{x})$$
$$+ r_i P(\mathbf{x}, \boldsymbol{\lambda}) \star R_{i+1}(\mathbf{x})], \tag{A12}$$

where

$$P(\mathbf{x}, \boldsymbol{\lambda}) * R_{i-1}(\mathbf{x}) = \int w\left(\mathbf{x} - \boldsymbol{\eta} + \frac{\boldsymbol{\lambda}}{2}\right) \cdot$$
$$P(\mathbf{x}, \boldsymbol{\lambda})\, R_{i-1}(\boldsymbol{\eta} - \boldsymbol{\lambda})\, d\boldsymbol{\lambda} \tag{A13}$$

and

$$P(\mathbf{x}, \boldsymbol{\lambda}) \star R_{i+1}(\mathbf{x}) = \int w\left(\mathbf{x} - \boldsymbol{\eta} - \frac{\boldsymbol{\lambda}}{2}\right) \cdot$$
$$P(\mathbf{x}, \boldsymbol{\lambda})\, R_{i+1}(\boldsymbol{\eta} + \boldsymbol{\lambda})\, d\boldsymbol{\lambda} \tag{A14}$$

are the local convolution and local correlation of the penalty operator with the image over the local window w, respectively. The derivation for the receiver-side adjoint source is totally analogous and leads to the formula:

$$\nabla_{u_{r,i}} \mathcal{J} = u_{s,i}(\boldsymbol{\eta}, \mathrm{t}) \left[r_{i-1} P(\mathbf{x}, \boldsymbol{\lambda}) * R_{i-1}(\mathbf{x})\right.$$
$$\left. +\ r_i P(\mathbf{x}, \boldsymbol{\lambda}) \star R_{i+1}(\mathbf{x})\right]. \tag{A15}$$

Equations (A12) and (A15) are the expressions for the adjoint sources in equations (14) and (15) in the paper.